Pergamon

PII: S0042-6989(96)00310-0

Vision Res., Vol. 37, No. 12, pp. 1595–1604, 1997

Printed in Great Britain
0042-6989/97 $17.00 + 0.00

Principles of an Adaptive Method for Measuring the Slope of the Psychometric Function

P. EWEN KING-SMITH,*‡ DAVID ROSE†

Received 28 November 1995; in revised form 26 August 1996

Recent developments in the efficient estimation of threshold are here extended to the problem of how best to estimate the slope of the psychometric function. An adaptive method is described for selecting stimulus intensities that are optimal for slope estimation. A two-dimensional array of probabilities of different thresholds and slopes is used to calculate the stimulus intensity for the next trial; this array is updated after the trial, using Bayes' theorem to incorporate information from the subject's response. The practical implementation and efficiency of the method are demonstrated and discussed. © 1997 Elsevier Science Ltd.

Psychophysics Psychometric function Method Threshold

INTRODUCTION

The psychometric function relates observer performance to the intensity of a stimulus. Many experiments in psychophysics seek to find the stimulus intensity or value that elicits some criterion level or threshold of performance on a task, and numerous papers have addressed the question of how this can be done most efficiently (see review by Treutwein, 1995). Efficient methods are those that enable the estimation of threshold, to within a specified limit of accuracy, in the fewest trials.

However, relatively little consideration has been given to the question of how to measure the other parameters of the psychometric function, particularly its slope (i.e., the rate of change of performance with stimulus intensity). Many methods for threshold estimation assume a particular value for the slope of the psychometric function, and that this value remains constant throughout the experiment. If the estimated slope is inaccurate, however, inefficiencies of threshold estimation can occur.§ Most importantly, however, the slope parameter can be of significance in its own right. Accurate estimation of the slope is crucial to a number of issues, from tests of basic sensory theory to applications in neurology. For example, in visual psychophysics it is necessary to measure the slope of the function relating stimulus contrast to detection in order to test the predictions of the high-threshold probability summation and uncertainty models of contrast detection (Nachmias, 1981; Pelli, 1985; Mayer & Tyler, 1986) and binocular summation (Legge, 1984), and the multi-stage differential coupling model of Laming (1986). In a luminance (or contrast) discrimination experiment (Nachmias & Sansbury, 1974; Whittle, 1986), discrimination ability is given by the slope of the function relating probability of choosing one of two stimuli as a function of the difference in luminances between the two stimuli. In the clinic, changes in the slope of the psychometric function for visual detection occur in multiple sclerosis, and assessment of the slope may thus be useful for diagnosis of the disease (Patterson *et al.*, 1980). More generally, the effects of drugs are often assessed in terms of the slope of the dose–response curve (Finney, 1978).

Several methods for slope estimation exist (e.g., Watt & Andrews, 1981; Hall, 1981; Levitt, 1971; Leek *et al.*, 1992). However, none take advantage of the most recent developments in the efficient threshold methods. These developments involve the use of on-line computers to keep running estimates of the probable values of the target parameter (e.g. threshold) throughout the experiment, and in between every trial they select the optimal value of stimulus to present on the next trial (e.g. Watson & Pelli, 1983; Shelton, 1983; King-Smith, 1984; Emerson, 1986; Harvey, 1986; King-Smith *et al.*, 1994). For maximum efficiency, the value chosen must be such that, given the expected probabilities of each of the possible responses that the subject might give, the variability of the estimation of threshold will be reduced as much as possible by the information obtained from the subject on that trial (or, ideally, over the course of all the following trials; Pelli, 1987). This stimulus value is now

*College of Optometry, The Ohio State University, 338 West Tenth Avenue, Columbus, OH 43210-1240, U.S.A.

†Department of Psychology, University of Surrey, Guildford, Surrey GU2 5XH, U.K.

‡To whom all correspondence should be addressed [*Fax* +1-614-292-7493; *Email* king-smith.1@osu.edu]

§In practice, the inaccuracy in estimated threshold may be minor (e.g., Green, 1993); however, the confidence limits of that estimate may be relatively inaccurate (King-Smith *et al.*, 1994).

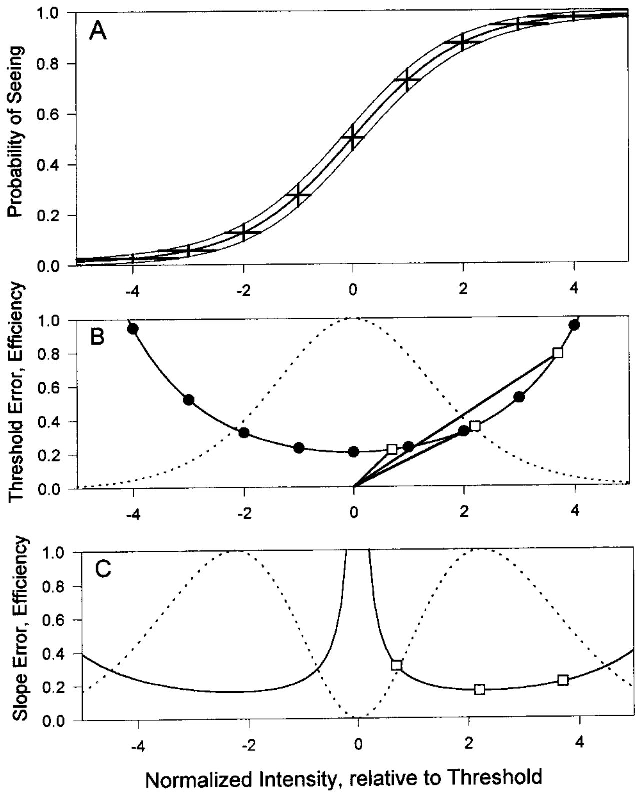

FIGURE 1. Derivation of "threshold efficiency" and "slope efficiency". (A) Thick curve is the psychometric function plotted as a function of "normalized intensity" [Eq. (2)]. Vertical bars correspond to $\pm$ 1 SE in probability for $n = 100$ trials; the thin curves have been drawn through the ends of these bars. The horizontal bars therefore give ± 1 SE in normalized intensity, for the corresponding measured probabilities; if the form and slope of the psychometric function are known, the horizontal bars correspond to the standard errors in the threshold estimate, based on 100 trials at the corresponding intensity. (B) The filled circles and solid curve give the standard error of the threshold estimate (in normalized units) derived in the manner indicated in (A). The dotted curve is the "threshold efficiency" which is inversely proportional to the variance of the threshold estimate and is scaled to a maximum of 1.0. The straight lines drawn to the open squares illustrate the derivation of errors in the slope of the psychometric function when the threshold is known; these errors are proportional to the gradients of these lines (see text). (C) Solid curves and squares give the standard error of slope estimates [see (B) and text]. The dotted curve gives the "slope efficiency" which is inversely proportional to the variance of the slope estimate and is scaled to a maximum of 1.0.

agreed to be, in most circumstances, close to the current estimate of the threshold (Harvey, 1986; King-Smith *et al.*, 1994).

Recently, it has been realized that the same principles that have led to the most efficient procedures for threshold estimation can also be applied to the problem of efficient slope estimation (Rose, 1988; King-Smith *et al.*, 1994, 1995). In this paper we describe the principles of this method and discuss its advantages, limitations and practical implementation. The problems in performing adaptive psychophysical experiments of this type are (1) deciding the a priori estimated probability distributions for the threshold and slope values; (2) deciding which stimulus values to present on each trial, and (3) analyzing the final data set to obtain the most accurate estimates of the true threshold and slope (King-Smith *et al.*, 1994). In this paper we will deal principally with issue (2), although implications for the other two problems will naturally arise.

GENERAL PRINCIPLES

Choice of intensities for threshold estimation

If the form and slope of the psychometric function are known, then threshold can be estimated from the probability of seeing at a certain intensity. For a given number of trials, some intensities yield more precise estimates of threshold than others. Errors associated with different intensities are illustrated in Fig. 1(A), where the thick solid curve represents a logistic psychometric function of the form

$$p = \gamma + (1 - \gamma - \delta)/(1 + \exp(-\chi)) \tag{1}$$

$$\chi = k(x - T) \tag{2}$$

where γ is the false positive rate (0.01 in this example), δ is the false negative rate (0.01), χ will be called normalized intensity, k is the "slope parameter" which determines the slope of the psychometric function, x is log intensity and T is log threshold. Normalized intensity, defined as above, is plotted on the abscissa in Fig. 1. Vertical bars in Fig. 1(A) represent ± 1 standard error (SE) of the estimated probability of seeing for $n = 100$ trials given by the binomial formula

$$E_p = (p(1 - p)/n)^{1/2} \tag{3}$$

where p is the predicted probability of seeing. ($n = 100$ is used for purposes of illustration, because it provides a satisfactory size of error bars in Fig. 1(A); it may be noted that the binomial distribution is asymmetrical when p is equal or close to 0 or 1, but in those cases it becomes more symmetrical as n is increased, which could be done without affecting the conclusions of this analysis). The thin curves have been drawn through the upper and lower ends of these bars. Because these curves represent ± 1 SE, it follows that for any measured probability, ± 1 SE, E_χ, of normalized intensity is given by the length of the corresponding horizontal bar in Fig. 1(A); these bars correspond to the standard error in log threshold [because the slope parameter, k, and log intensity, x are known in Eq. (2)]. Threshold errors, derived using this principle from the formula

$$E_\chi = E_p/(\mathrm{d}p/\mathrm{d}\chi) \tag{4}$$

(Taylor & Creelman, 1967: Taylor, 1971; Green *et al.*, 1989) have been plotted in Fig. 1(B) (circles and solid curve); Eq. (4) can be derived by assuming that the psychometric function is locally linear (which is a good approximation when n is large and so the error bars are small). It is seen that standard error is minimal for a normalized intensity of 0 corresponding to 50% probability of seeing. This will be true for any value of n. Efficiency (Taylor & Creelman, 1967) is defined to be inversely proportional to variance with a maximum of 1 and is shown by the dotted curve of Fig. 1(B) (cf. Levitt, 1971); we will call this "threshold efficiency" to distinguish it from "slope efficiency" discussed below.

Choice of intensities for slope estimation

Now suppose that threshold intensity (corresponding to $p = 0.5$) is known whereas the slope parameter, k, is to be estimated by measuring the probability of seeing at a given intensity. As before, the standard error, E_χ, in normalized intensity, $\chi = k(x-T)$, is given by the solid curve in Fig. 1(B) and this error is now due to the standard error in the slope parameter, k (because both x, log intensity, and T, log threshold, are known). The fractional error in the slope estimate is thus equal to the fractional error in normalized intensity, i.e.,

$$E_k/k = E_\chi/\chi \tag{5}$$

and so is given by the gradient of a line joining the origin to the corresponding point on the threshold error curve in Fig. 1(B), which plots E_χ as a function of χ. Three examples are given by the lines drawn to the open squares in Fig. 1(B); it is seen that the fractional error in the slope parameter, E_k/k, is least for a normalized intensity of about 2.2. These "slope errors" [Eq. (5)] are plotted as the squares and solid curve in Fig. 1(C). "Slope efficiency" (i.e., efficiency of estimating the slope parameter) can now be defined as being inversely proportional to the variance of the estimate of slope parameter, k, with a maximum of 1, and is given by the dotted curve in Fig. 1(C) (cf. Levitt, 1971).

The preceding analysis uses the implausible assumption that the 50% threshold is known so accurately that the error in the slope parameter is due only to the one additional measurement at a normalized intensity near 2.2. However, it may be shown that, for a symmetrical psychometric function such as that in Fig. 1(A), a similar analysis applies to the more realistic strategy of equal numbers of trials at two log intensities equally above and below log threshold (Wetherill, 1963; O'Regan & Humbert, 1989). Thus, maximum slope efficiency could be obtained by using normalized intensities of ± 2.24 [i.e., at the peak slope efficiencies in Fig. 1(C)] corresponding to probabilities of seeing, p, of 0.896 and 0.104; the corresponding threshold efficiency from Fig. 1(B) would be 0.324.

Conclusions re: choice of intensities

In conclusion, greatest threshold efficiency is obtained by using intensities near threshold ($p = 0.5$). Greatest slope efficiency is obtained by using equal numbers of trials near two log intensities symmetrically above and below log threshold.* Importantly, both threshold and slope efficiencies are independent of the number of trials, n, and the number of standard errors used for the error bars [which was 1.0 in Fig. 1(A)]; (however, it should be emphasized that precision, which is the reciprocal of variance, does depend on n and increases with increasing n.) Comparison of Fig. 1(B) and Fig. 1(C) shows the possible trade-off between threshold and slope efficiencies; for example if the spacing between the two log intensities is reduced, threshold efficiency increases at the expense of reduced slope efficiency. In the simulations and experiment described below, the computer has attempted to set normalized intensities of ± 2.07 (rather than the optimal ± 2.24), corresponding to $p = 0.880$ and 0.120; for these conditions, slope efficiency is still near maximum (0.990), while threshold efficiency is moderately increased to 0.376.

It should be emphasized again that the preceding discussion is based on the assumptions that both threshold and slope are fairly well known before the experiment, so that optimum intensities can be chosen with the above criteria; in practice, this is unlikely to be true, so the adaptive method of this article has been developed, using the above principles, but for the normal situation where threshold and slope are less well known before the experiment.

Estimation of threshold: the standard ZEST method

Before describing an adaptive Bayesian method for measuring both threshold and slope, it is helpful to review how Bayes' theorem can be applied to measuring threshold (while assuming a constant slope parameter). The principles of efficient estimation of threshold are illustrated in Fig. 2, which illustrates the first trial of a yes–no experiment. Figure 2(A) represents the experimenter's knowledge or guess about relative probabilities of different threshold values being the "true" threshold. In this case log threshold = 0 is guessed to be the most probable value, and the assumed probability falls off as a hyperbolic secant of log threshold. (Another bell-shaped curve such as a gaussian distribution could be used; we prefer the hyperbolic secant because it does not decay so rapidly to zero at low and high log thresholds, which results in more efficient measurement of thresholds which are far from the best guess). Log intensity for the first trial is set to the center of this function, as indicated in Fig. 2(A). A response of "yes" (or "no") will provide further information about the probabilities that each of the different candidate log thresholds is the true one. This information can be expressed as a likelihood function [Fig. 2(B)] which is the probability of a "yes" (or "no") response to a stimulus of this intensity as a function of the subject's log threshold. The "yes" function is a left-to-right mirror image of the psychometric function and so its

*Somewhat different rules apply to an asymmetrical psychometric function, e.g., a Weibull function or a forced choice experiment (Watson & Pelli, 1983; O'Regan & Humbert, 1989). Our analysis shows that, for optimal estimation of slope, it may be theoretically preferable to present relatively more trials at one of the two intensities; for example, for a yes–no Weibull function, more trials should be presented at the lower intensity, whereas, for a two-alternative forced choice experiment and a logistic function, more trials should be presented at the higher intensity. The current method could readily be modified to present more trials at the corresponding intensity. We are developing analytical and graphical methods for choosing optimal intensities and relative numbers of trials for these asymmetrical psychometric functions (manuscript in preparation).

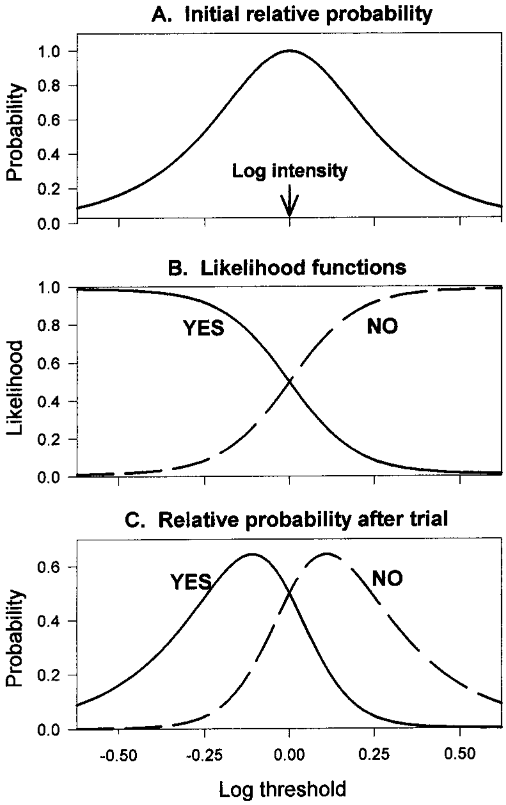

FIGURE 2. First trial of the standard ZEST method (King-Smith *et al.*, 1994). (A) The experimenter's prior knowledge of threshold, i.e., the probability of different log threshold values in the experimental population and conditions. The log intensity chosen for the first trial is marked. (B) Likelihood functions for "yes" and "no" responses to the first stimulus; for example, the "yes" likelihood function is the probability that the subject responds "yes" to a trial of this intensity as a function of log threshold. (C) Probabilities of different threshold values after the first trial—by Bayes' theorem, these curves are proportional to the product of the function in (A) with the corresponding function in (B).

slope (and that of the "no" function) depends on the assumed slope parameter of the psychometric function.*

The information in Fig. 2(A, B) can be combined by "Bayesian multiplication" [i.e., multiplication of the initial probability of Fig. 2(A) with a likelihood function of Fig. 2(B)] to describe the relative probabilities of different log threshold values after the trial [Fig. 2(C)]. (Note that the functions describing the probabilities of different values of log threshold, as in Fig. 2(A) and Fig. 2(C) are often expressed as "probability density functions", where the ordinate is scaled so that the area under the curve is 1.0. This scaling has not been performed in Figs, 2–5, because it is unnecessary for implementing the current method; we will call these curves "relative probability functions".)

For the second trial, the relative probability function in Fig. 2(C) is placed into the top panel and is used to choose the new stimulus intensity. Note that the distribution is no longer symmetrical, so we have to decide which log intensity is best (e.g., the mode? the mean?). It turns out that the most efficient stimulus to use is very close to the mean of the distribution (King-Smith *et al.*, 1994), so this is the value that is presented on trial 2: it is -0.16 if the subject responded "yes" on the first trial, or $+0.16$ if he responded "no". The likelihood functions in the middle panels now have to be slid along by the same amount (-0.16 or $+0.16$ log units, respectively) so they are centered on the presented stimulus value. The top and middle panels are then combined again according to the subject's "yes" or "no" response on trial 2.

This process can be repeated many times using the relative probability function after one trial as the initial probability for the next trial. At the end of the experiment, the relative probability function that remains will generally be much narrower than the initial function shown in Fig. 2(A). Finally, an estimate of log threshold is given by the mean of the final relative probability function (King-Smith, 1984; Emerson, 1986; King-Smith *et al.*, 1994).

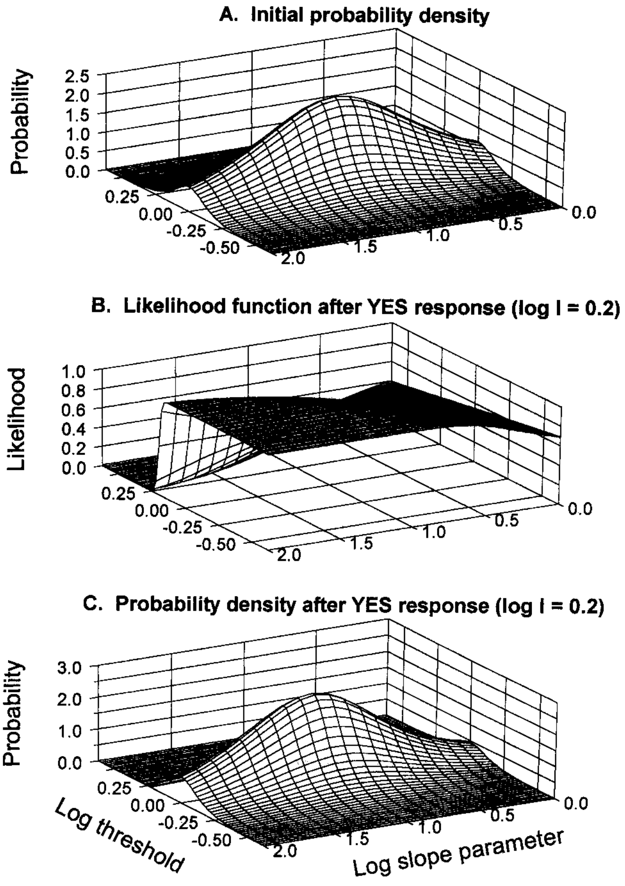

FIGURE 3. First trial of the modified ZEST method. (A) The experimenter's prior knowledge of log threshold and log slope. (B) Likelihood function for a "yes" response; in this example, the log intensity of the trial was 0.2 corresponding to an estimated probability of seeing of 0.88. This is an "expected response" to a "high" intensity (see text). (C) Probabilities of different log threshold and log slope values after the first trial—this function is proportional to the product of the functions in (A) and (B).

*The psychometric function is the logistic function of Eq. (1), but any reasonable shape can be used, including a Weibull or a cumulative gaussian function.

Simultaneous estimation of slope and threshold

Whereas efficient estimation of threshold requires all test intensities to be placed close to the current estimate of threshold, efficient estimation of the slope parameter requires that roughly equal numbers of trials be placed at two intensities above and below threshold (Fig. 1), which will be described as "high" and "low" intensities, respectively. The term "expected response" will be used for responses with high expected probabilities, i.e., a "yes" response to a high intensity or a "no" response to a low intensity; the term "unexpected response" will mean the opposite, i.e., a "no" response to a high intensity or a "yes" response to a low intensity.

Each trial is chosen to be at either a high or a low intensity, at random. Figure 3 illustrates the computations involved for a "yes" response on the first trial of an experimental run which starts with a high intensity (estimated $p = 0.88$); it is thus an expected response. This figure is analogous to Fig. 2 (yes response), except that it shows functions of two variables, log threshold (T) and log slope parameter (log k). Figure 3(A) represents the experimenter's knowledge or guess about the probability of different values of log threshold and log slope parameter. This bell-shaped function is a product of hyperbolic secants of log threshold and log slope (other bell-shaped functions such as gaussians could be used—we prefer the hyperbolic secant for reasons discussed previously). The bell-shaped function is relatively narrow along the log threshold dimension and broad along the log slope dimension, corresponding to the experimental example (luminance discrimination) to be described below (Fig. 5); in a detection threshold experiment, the initial relative probability function would probably be broader along the log threshold dimension than along the log slope dimension. Estimated log threshold and log slope can be derived from the center of gravity of the relative probability function in Fig. 3(A), $q(T, \log k)$, by the formulae

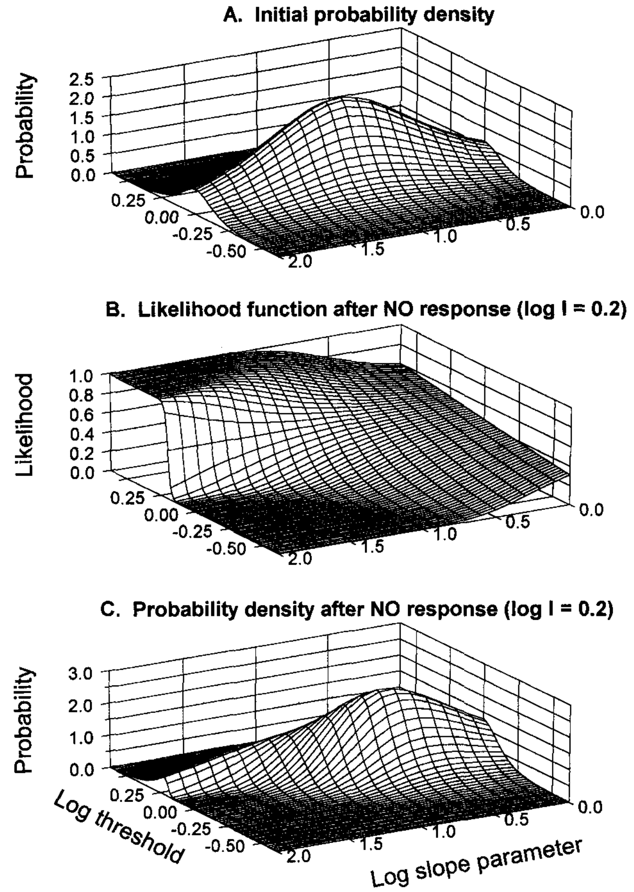

FIGURE 4. First trial of the modified ZEST method, as in Fig. 3, but for a "no" response; this is an "unexpected response".

Figure 3(B) illustrates the likelihood function for a "yes" response at the log intensity, x, of 0.20. This likelihood function can be considered to be a family of "yes" likelihood functions, like that of Fig. 2(B), each with a different value of log slope parameter; note that log threshold = 0.20 corresponds to the intensity used in

$$\text{Estimated log threshold} = E_T = \Sigma\Sigma T q(T, \log k)/\Sigma\Sigma q(T, \log k) \tag{6}$$

$$\text{Estimated log slope} = E_{\log k} = \Sigma\Sigma(\log k) q(T, \log k)/\Sigma\Sigma q(T, \log k) \tag{7}$$

where the double summation is over the ranges of values used for T and log k. The estimated log threshold is $T = 0$, and log slope parameter, $\log k = 1$, (thus the slope parameter, $k = 10$). The estimated probability of seeing of 0.88 corresponds to a normalized intensity, χ, of 2.07. Thus, log intensity for the first trial can be determined by rewriting Eq. (2)

$$x = T + \chi/k \tag{8}$$

and so is 0.207; in practice, it is convenient to vary log intensity in discrete steps of, say, 0.02 log units, so rounded to the nearest step, log intensity becomes 0.20. For a low intensity trial with an estimated probability of 0.12, normalized intensity would be -2.07, leading to a log intensity of -0.20.

this trial, so that the likelihood of this threshold is 0.5, independent of slope parameter [Eq. (1)]. The resultant probability after the first trial is given by Fig. 3(C); as in Fig. 2, this is the Bayesian product of upper and middle panels. The relative probability function in Fig. 3(C) is used as the initial relative probability for the next trial; in this case the "yes" response in the first trial has reduced estimated log threshold slightly from 0 to -0.028. Estimated log slope has increased slightly from 1 to 1.0654 (this increase is a consequence of the fact that the "yes" response at the "high" intensity is an "expected" response). The next intensity, which can be chosen at random to be either high or low, can now be calculated using Eq. (8). The process illustrated in Fig. 3 can be repeated for as many trials as necessary.

Figure 4 illustrates the calculations involved when the first response to the same high intensity (as in Fig. 3) is "no" and so is an unexpected response. The "no" likelihood function in Fig. 4(B) is a mirror image of the "yes" likelihood function in Fig. 3(B) [cf. "yes" and "no" likelihood functions in Fig. 2(B)]. The resultant relative probability is shown in Fig. 4(C). As can be seen from this plot, this "no" response has increased the estimated log threshold considerably, from 0 to 0.099, and the estimated log slope has been reduced considerably, from 1 to 0.770—a consequence of the fact that the response was unexpected.

AN EXPERIMENTAL EXAMPLE

Threshold and slope estimation

Figure 5 illustrates an actual threshold and slope determination using 50 trials. The subject's task was to indicate which of two spots appeared brighter—a "yes" response indicating that the upper spot appeared brighter than the lower spot. The relative luminance of the two spots was varied while keeping their mean log luminance constant; "log intensity" was defined as the logarithm of the ratio of upper to lower luminance. In this case "log threshold" corresponds to the point of subjective equality between the two spots and hence should be zero for an ideal observer with no asymmetry between upper and lower visual field (in fact, the current method was developed because observers were found to be not ideal in this respect); "log slope" is a measure of luminance discrimination between the spots when they differ from subjective equality. As before, the upper panel shows the experimenter's prior knowledge of slope parameter and threshold. This function, which is given by

$$\mathrm{sech}(10.T).\mathrm{sech}(2((\log k) - 1))$$

was chosen on the basis of some preliminary experiments (see also "Concluding Remarks"); it is relatively narrow along the log threshold (T) axis because large asymmetries between upper and lower fields were not expected; however, the function is relatively broad along the log slope parameter ($\log k$) axis, because discrimination of log luminance varies considerably in different experimental conditions, e.g., different values of mean log luminance (Whittle, 1986). The likelihood function in Fig. 5(B) is the product of the likelihood functions for all 50 trials, and Fig. 5(C) is, as before, the product of the upper and middle panels. The information collected from the subject has narrowed the uncertainty from that of the initial estimation. It can be seen that log threshold has been determined relatively accurately, while log slope parameter has not been determined quite so accurately.

Final estimates of log threshold and log slope parameter were calculated from Eqs (6) and (7), where $q(T, \log k)$ now refers to the final relative probability function of Fig. 5(C). Standard errors of these estimates were calculated from the variance of the final probability function along T and $\log k$ axes; thus

$$\text{SE in log threshold} = \sqrt{}[\Sigma\Sigma(T - E_T)^2 q(T, \log k)/\Sigma\Sigma q(T, \log k)]$$

$$\text{SE in log slope parameter} = \sqrt{}[\Sigma\Sigma(\log k - E_{\log k})^2 q(T, \log k)/\Sigma\Sigma q(T, \log k)]$$

where E_T and $E_{\log k}$ are the estimated values of log threshold and log slope parameter [Eqs (6) and (7)]. The estimated log slope parameter, $E_{\log k}$, was found to be 1.324 ± 0.125 SE (estimated $k = 21.1$); estimated log threshold, E_T, was -0.104 ± 0.023 SE. Both parameters thus differ significantly from the initial guesses ($\log k = 1$; $T = 0$).

The two-dimensional relative probability in Fig. 5(C) has been converted to a relative probability of log slope parameter by integrating over log threshold; the corresponding probabilities are plotted on a logarithmic scale as the circles in Fig. 6(A). The solid curve represents a gaussian function with the same mean (1.324) and standard deviation (0.125). The dotted curve gives the initial relative probability scaled to the same maximum for easier comparison; the final probability function is much narrower than the initial one, so that information about log slope is derived mainly from the experimental data, rather than from the initial assumptions. It is seen that the final probability is fairly well fitted by the gaussian function for probabilities above about 1% of maximum. The circles and solid curve in Fig. 6(B) give the corresponding cumulative probabilities (i.e., the probability that log slope parameter will be less than the value on the abscissa); a "normal probability scale" has been used, so that a gaussian function plots as a diagonal straight line. Dashed lines have been drawn at 2.5 and 97.5% probabilities, thus defining the 95% confidence interval. It is seen that, over this interval, the calculated cumulative probability (circles) agrees well with the gaussian (solid line) indicating that the confidence range can be predicted quite well from the gaussian. However, outside this interval, the two cumulative probabilities deviate, so it would be preferable to use the calculated cumulative probability rather than the gaussian for deriving wider confidence intervals (e.g. 99% or 99.9%).

To illustrate further the empirical procedure, the intensities presented on all 50 trials are shown in Fig. 7(A). Open symbols correspond to "yes" responses (upper spot appears brighter than lower) whereas filled symbols represent "no" responses. Log intensities above -0.1 correspond to "high" intensities (expected probability of response of 0.88) whereas log intensities below -0.1 correspond to "low" intensities (expected $p = 0.12$). Circles and triangles represent expected and unexpected responses, respectively. Estimated log thresholds and log slopes before each trial are represented

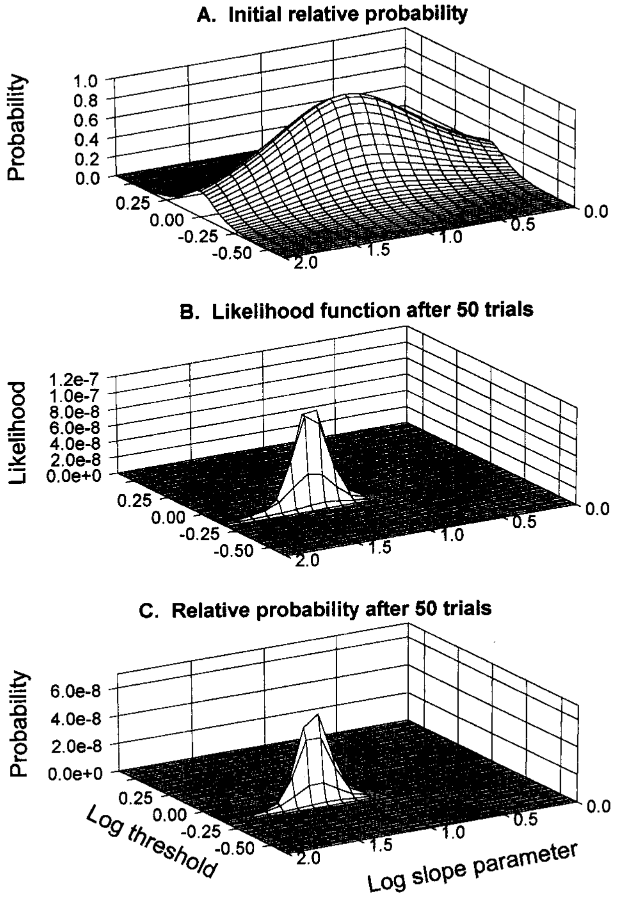

FIGURE 5. Results obtained from an experimental run of 50 trials, using the modified ZEST method; the subject's task was to indicate which of two test spots appeared brighter, a "yes" response meaning that the upper spot appeared brighter than the lower spot. In this experiment, log intensity corresponds to the logarithm of the ratio of upper to lower luminances, and log threshold is the corresponding value of log intensity which gives a 50% probability of a "yes" (upper) response. See text and caption to Fig. 3 for details.

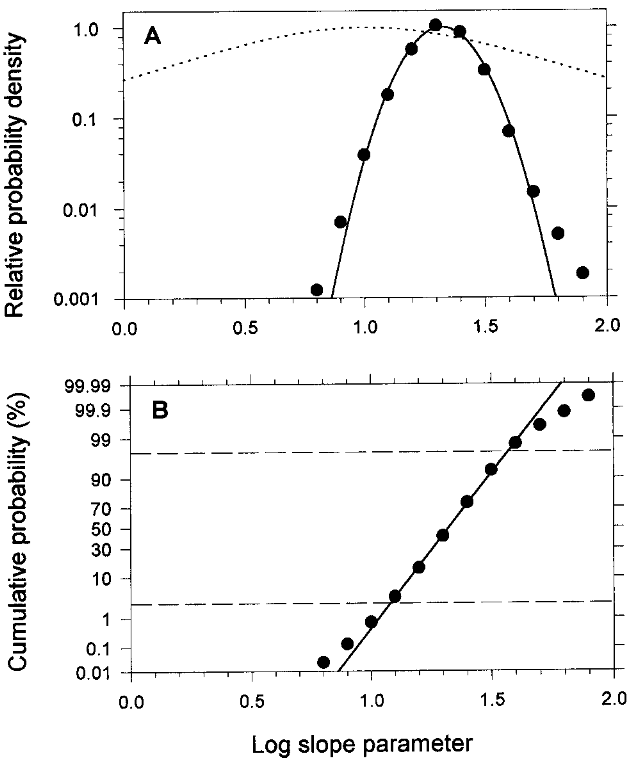

FIGURE 6. (A) Filled circles give the probability of different log slope values at the end of the experiment; this is obtained by integrating the final probability function in Fig. 5(C) over log threshold. The solid curve is a gaussian function of the same mean and standard deviation. The dotted curve gives the initial probability [derived from Fig. 5(A)]. A logarithmic scale of probability has been used. (B) Cumulative probabilities—i.e., the probability that log slope is less than the value on the abscissa. Circles are the calculated probabilities derived by integrating the corresponding data in (A), using Simpson's rule (Press *et al.*, 1992) and a step size of 0.05 log slope units; likewise, the solid line is derived from the gaussian in (A). Dashed lines indicate cumulative probabilities of 2.5% and 97.5%, thus defining a 95% confidence interval. A "normal probability scale" is used so that the gaussian yields a straight line.

in Fig. 7(B) and Fig. 7(C) respectively; squares represent final estimates. These plots illustrate the following aspects of the method:

1. "yes" responses lower estimated log threshold, "no" responses raise estimated log threshold [Fig. 7(B)].
2. Expected responses raise estimated log slope, unexpected responses lower estimated log slope [Fig. 7(C)].
3. Expected responses cause small changes (in both log threshold and log slope), unexpected responses cause larger changes.
4. Changes (in both log threshold and log slope) tend to be large at the beginning of the experimental run and become smaller as the run progresses.

The subject's estimated probability of "yes" (upper spot appears brighter) response curve is shown in Fig. 8(A), using the estimated value of log threshold (−0.104) and log slope parameter ($\log k = 1.324$, $k = 21.1$). The vertical dashed line corresponds to physical equality of the two spots, whereas the vertical dotted line corresponds to subjective equality (50% "yes" or "upper" responses). It is clear that the upper spot tends to appear brighter than the lower spot, so that about 90% of the responses are "yes" (upper) when the two spots are physically equal (we checked that this was not due to equipment artifacts). Given this strong bias in the subject's responses, it is evident that a standard two-alternative forced-choice method, which assumes that subjective equality corresponds to objective equality, would be unsuitable for measuring intensity discrimination in this subject.

Actual efficiency

A histogram of the "log intensities" used in this experimental run is given in Fig. 8(B); high and low intensity trials form two distinct sub-histograms in this example. "Yes" and "no" responses are represented by white and black areas, respectively; thus there were two unexpected "no" responses at high intensities and five unexpected "yes" responses at low intensities. The theoretical analysis of Fig. 1 can now be applied to these

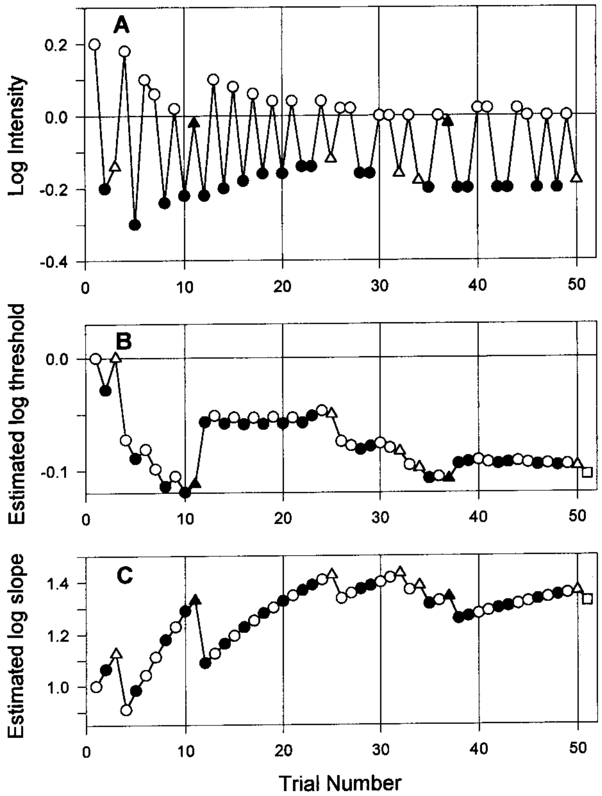

FIGURE 7. (A) Log intensities (ratios of upper to lower luminances) used for each trial of the experimental run. Open circles represent "yes" responses (upper spot appears brighter), whereas filled circles represent "no" responses. Points above a log intensity of −0.1 correspond to "high" intensities with an estimated probability of seeing, $p = 0.88$, whereas points below −0.1 correspond to "low" intensities ($p = 0.12$). Expected and unexpected responses are given by circles and triangles, respectively. (B) Estimated log thresholds before each trial. (C) Estimated log slope parameter before each trial. Squares in (B) and (C) indicate final estimates.

data. The solid line in Fig. 8(C) corresponds to estimated threshold efficiency [cf. Fig. 1(B)], whereas the dotted line corresponds to estimated slope efficiency [cf. Fig. 1(C)]. The two sub-histograms peak at log intensities fairly near the maxima for slope efficiency, implying fairly efficient estimation of slope; using Taylor's (Taylor, 1971) method of calculating "total precision gained" in the run, slope efficiency was calculated to be 58%. The trial intensities are rather poorly distributed relative to the single peak in threshold efficiency; correspondingly, threshold efficiency was calculated to be 31% and this relatively low efficiency is to be expected for a method which was optimized for slope measurements, rather than for threshold measurements.

PRACTICAL IMPLEMENTATION

The experimental run illustrated in Figs 5–8 was implemented using machine code subroutines on a North Star Horizon computer (4 MHz, 8 bit, Z80 microprocessor with a floating point coprocessor). Before the experiment, the initial probability function [Fig. 5(A)] was calculated as an array of 9 log slope parameter values (from 0 to 2 in steps of 0.25) by 63 log threshold values (from −0.62 to 0.62 in steps of 0.02). These ranges were chosen to cover the expected range of log slope and log threshold, with margins at the ends of the ranges, so that most of the final relative probability functions [e.g. Figure 5(C)] would always be included. Similarly, "yes" and "no" likelihood functions (as in the center panels of Figs 3 and 4) were calculated as arrays of the same 9 log slope values by 125 values of log threshold minus log intensity (from −1.24 to 1.24 in steps of 0.02); the wider range of log threshold used for the likelihood function permits the "sliding" along the log threshold axis described below.

During the experiment, the calculated log intensity for the next trial was rounded to the same step size, i.e., to the nearest 0.02. Therefore, the Bayesian multiplication (e.g., in Fig. 3) may be performed by sliding the likelihood function array along the log threshold axis by an amount given by the log intensity of the stimulus, before multiplying the current probability array; in this way, it is not necessary to perform the lengthy calculations of a new likelihood array for each trial (Watson & Pelli, 1983; Shelton, 1983). However, with the increased speed of modern computers, it may be possible to calculate the two-dimensional likelihood array in real time between trials; this would be advantageous when only a limited set of stimulus intensities is available, which does not match the uniform step in log intensity of a pre-calculated likelihood array (Harvey, 1986).

The size of the probability array (567 probability values) in our experiment was limited by the processing speed of the microcomputer, which took some 3 sec to perform the Bayesian multiplication for the current response, and to calculate the log intensity for the next stimulus. We have found empirically that the step size in log slope parameter and log threshold values should ideally be no greater than the corresponding standard errors in the final estimates. In this respect, our choice of step in log threshold (0.02) was satisfactory compared to the standard error of the log threshold estimate (0.023). However, the step size in log slope parameter (0.25) was greater than ideal when compared to the standard error of the estimate of log slope parameter (0.125); a more accurate estimate of log slope parameter was performed after the experiment by repeating the Bayesian multiplications with smaller steps in log slope parameter (0.1, as in Figs 3–5) which showed that the original estimate of log slope parameter, 1.314 ± 0.130 should have been 1.324 ± 0.125. Further reducing the step size for either log threshold (to 0.01) or log slope parameter (to 0.05) did not change either the mean or standard error (to within three decimal places). Empirically, we find that, for calculating confidence intervals as in Fig. 6(B), smaller steps in log slope parameter are needed; for three decimal place accuracy, steps of 0.025 and 0.05 were sufficient using the trapezoidal rule and Simpson's rule, respectively (Press *et al.*, 1992). Because modern microcomputers are about 100 times faster than our system, it should be possible to perform the real-time

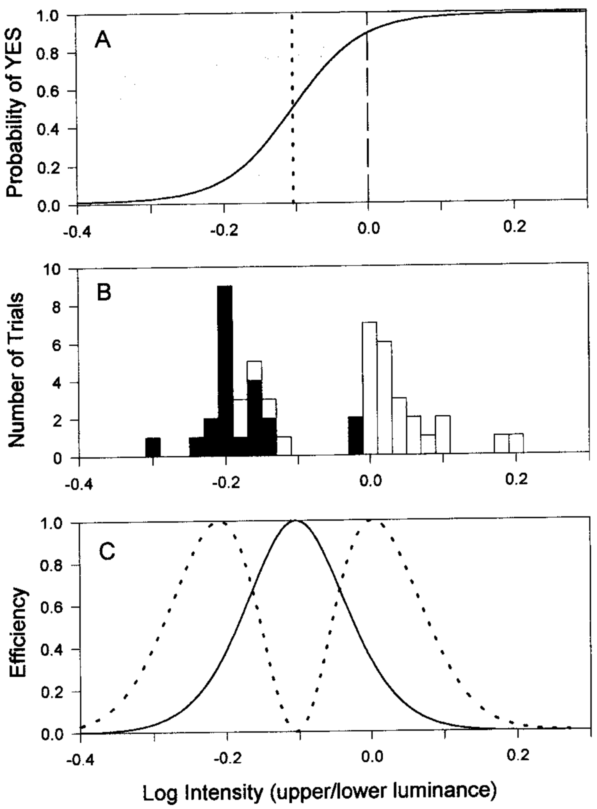

FIGURE 8. (A) The estimated probability of seeing curve derived from the experiment of Figs 5–7. The vertical dashed line corresponds to physical equality between the upper and lower spots, whereas the dotted line corresponds to subjective equality ($p = 0.5$). (B) Histogram of the distribution of log intensities for the 50 trials of the experimental run. White and black areas represent "yes" and "no" responses, respectively. (C) The dashed and dotted curves give estimated threshold and slope efficiencies, respectively (cf. Fig. 1); these are derived from the estimated values of log threshold and log slope.

calculations on much larger arrays than those reported here.

For very long experimental runs, it is theoretically possible for all the relative probabilities [e.g. as in Fig. 5(C)] to become smaller than the smallest positive number handled by the computer, in which case all these probabilities are set to zero and the program fails. For our computer, where the smallest positive number is 10^{-64}, this would occur for a run of about 400 trials. For such unusually long runs, this problem could be avoided by calculating probability density functions rather than relative probabilities, i.e., by scaling the ordinate in Fig. 3(C) etc., after each trial, so that the volume under the surface equals 1.0.

The experiment can be run until it reaches a stopping criterion (e.g., the standard error in log slope parameter reaches a preset limit), or else it can be run for a fixed number of trials (as in Fig. 5), which eliminates the need to check between every trial whether the stopping criterion has been reached. Subjects typically prefer using a fixed number of trials so that the length of the experiment is well defined, rather than the indefinite length required to reach another stopping criterion; using a fixed number of trials may also avoid expending too much testing time on bad sessions when the subject is performing sub-optimally. The main parameters derived at the end of the experiment are the means and standard errors of log threshold and log slope parameter. However, it should be noted that it is also possible to calculate confidence intervals by constructing cumulative probability functions, as in Fig. 6(B).

CONCLUDING REMARKS

The standard ZEST method can be fine-tuned to increase efficiency and reduce bias, by using an initial relative probability function which matches the distribution of log thresholds found in the population under study (King-Smith *et al.*, 1994). Similar improvements might be obtained for the modified ZEST method which is described here, by analyzing log slope and log threshold data obtained in typical conditions. For intensity discrimination experiments such as Fig. 5, we have analyzed preliminary data from 60 experimental runs from five normal subjects in different experimental conditions. Estimates of log slope varied from 0.37 to 1.73 with a mean of 1.24 ± 0.30 (SD); this compares with the assumed initial relative probability function [Fig. 5(A)] which had a mean of 1.0 ± 0.48 (SD). Thus, for experiments using similar conditions, efficiency might be improved and bias might be reduced by increasing the mean value of log slope of the initial relative probability function and reducing its standard deviation. A more detailed analysis might indicate a skewed distribution which could be modeled with a modified hyperbolic secant (King-Smith *et al.*, 1994). With a faster computer, it would have been advisable to use a wider range of log slope in the calculations, e.g. from -0.5 to 2.5 rather than from 0 to 2 (as in Fig. 5); this would help ensure that all estimates and their corresponding relative probability functions [e.g., Fig. 5(C)] would lie well within the calculated range. Estimates of log threshold ranged from -0.269 to 0.192 with a mean of -0.034 ± 0.079 (SD); this compares to the initial relative probability function [Fig. 5(A)] of 0 ± 0.153 (SD). Again efficiency might be improved and bias reduced by using a rather narrower initial relative probability function. The range of log threshold used for calculation, -0.62 to 0.62, easily covered the range of estimated log threshold and was therefore probably adequate.

The experimental example of this paper, the intensity discrimination measurement of Figs. 5–8, is one in which slope, rather than threshold, is the main parameter of interest and it is poorly known prior to the experiment [Fig. 5(A)]. In these cases, threshold, which corresponds to deviation of the point of subjective equality from objective equality, is typically a "nuisance" parameter which is of less interest, and it is relatively well known before the experiment. By using equal numbers of "high" and "low" intensities to obtain relatively high estimated slope efficiencies but poor threshold efficiencies (Fig. 8), the procedure was tailored to favor the accurate estimation of slope. [Note that it is typically more difficult to measure slope accurately than to measure threshold, so that, even using conditions which favored estimation of slope, the final standard error of log slope parameter was greater than that of log threshold, as is seen in Fig. 5(C).]

An important modification of the present method might be for the case when threshold is the main parameter of interest, but slope is a nuisance parameter which may vary from subject to subject. In this case, two-dimensional probability arrays, as in Figs 3–5, would again be calculated as in those examples; however, intensity could be chosen on each trial at (or close to) the single value which optimizes threshold efficiency (rather than the two values which optimize slope efficiency). In this way, the precision of threshold estimation should be higher than in the current experiment, but the precision of slope estimation would be relatively poor. Thus, thresholds could be estimated efficiently, without the restrictive assumption made by ZEST (King-Smith *et al.*, 1994) that the slope parameter must have a certain value. This is similar to a proposal of modifying QUEST to maintain relative probability functions for multiple values of slope parameter, made by Watson and Pelli (1983). Estimates of the mean and standard error of the slope parameter within the population (e.g. King-Smith & Pierce, 1994), could be used in the initial relative probability function.

More experimental and theoretical work is needed to evaluate and quantify these proposals. A comparison of the efficiency of different methods of measuring slope, using Monte Carlo simulations, would be particularly valuable.

REFERENCES

Emerson, P. L. (1986). Observations on maximum-likelihood and Bayesian methods of forced-choice sequential threshold estimation. *Perception and Psychophysics, 39*, 151–153.

Finney, D. J. (1978). *Statistical method in biological assay*, 3rd edn. London: Charles Griffin & Co.

Green, D. M. (1993). A maximum-likelihood method for estimating thresholds in a yes–no task. *Journal of the Acoustical Society of America, 93*, 2096–2105.

Green, D. M., Richards, V. M. & Forrest, T. G. (1989). Stimulus step size and heterogeneous stimulus conditions in adaptive psychophysics. *Journal of the Acoustical Society of America, 86*, 629–636.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America, 69*, 1763–1769.

Harvey, L. O.Jr. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments, and Computers, 18*, 632–632.

King-Smith, P. E. (1984). Efficient threshold estimates from yes-no procedures using few (about 10) trials. *American Journal of Optometry and Physiological Optics, 61*, 119P.

King-Smith, P. E., Earley, M. J., Pierce, G. E. & Preston, J. L. (1995). An adaptive method for measuring both threshold and slope of the psychometric function. *Investigative Ophthalmology and Visual Science, 36*, S906.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C. & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research, 34*, 885–912.

King-Smith, P. E. & Pierce, G. E. (1994). Unbiased estimate of the slope of the psychometric function. *Investigative Ophthalmology and Visual Science* (Supp.), *35*, 1259.

Laming, D. (1986). *Sensory analysis*. London: Academic Press.

Leek, M. R., Hanna, T. E. & Marshall, L. (1992). Estimation of psychometric functions from adaptive tracking procedures. *Perception and Psychophysics, 51*, 247–256.

Legge, G. E. (1984). Binocular contrast summation—I. Detection and discrimination. *Vision Research, 24*, 373–383.

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.

Mayer, M. J. & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America A, 3*, 1166–1172.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21*, 215–223.

Nachmias, J. & Sansbury, R. V. (1974). Grating contrast discrimination may be better than detection. *Vision Research, 14*, 1039–1042.

O'Regan, J. K. & Humbert, R. (1989). Estimating psychometric functions in forced-choice situations: significant biases found in threshold and slope estimations when small samples are used. *Perception and Psychophysics, 46*, 434–442.

Patterson, V. H., Foster, D. H. & Heron, J. R. (1980). Variability of visual threshold in multiple sclerosis. *Brain, 103*, 139–147.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*, 1508–1532.

Pelli, D. G. (1987). The ideal psychometric procedure. *Investigative Ophthalmology and Visual Science* (Supp.), *28*, 366.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical recipes in C*, 2nd edn (pp. 129–132). Cambridge: Cambridge University Press.

Rose, D. (1988). Estimation of the slope of the psychometric function. *Perception, 17*, 396.

Shelton, B. R. (1983). Rapid calculation procedures for the maximum likelihood method in adaptive psychophysics. *Behavior Research Methods and Instrumentation, 15*, 87–88.

Taylor, M. M. (1971). On the efficiency of psychophysical measurement. *Journal of the Acoustical Society of America, 49*, 505–508.

Taylor, M. M. & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America, 41*, 782–787.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*, 2503–2522.

Watson, A. B. & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychophysical method. *Perception and Psychophysics, 33*, 113–120.

Watt, R. J. & Andrews, D. P. (1981). APE: adaptive probit estimation of psychometric functions. *Current Psychological Reviews, 1*, 205–214.

Wetherill, G. B. (1963). Sequential estimation of quantal response curves. *Journal of the Royal Statistics Society, B25*, 1–48.

Whittle, P. (1986). Increments and decrements: luminance discrimination. *Vision Research, 26*, 1677–1691.

Acknowledgements—The authors would like to thank Dr Bernhard Treutwein and an anonymous referee for their comments and suggestions. This work was supported by National Eye Institute Grant RO1 EY 04948 and the Ohio Lions Eye Research Foundation.